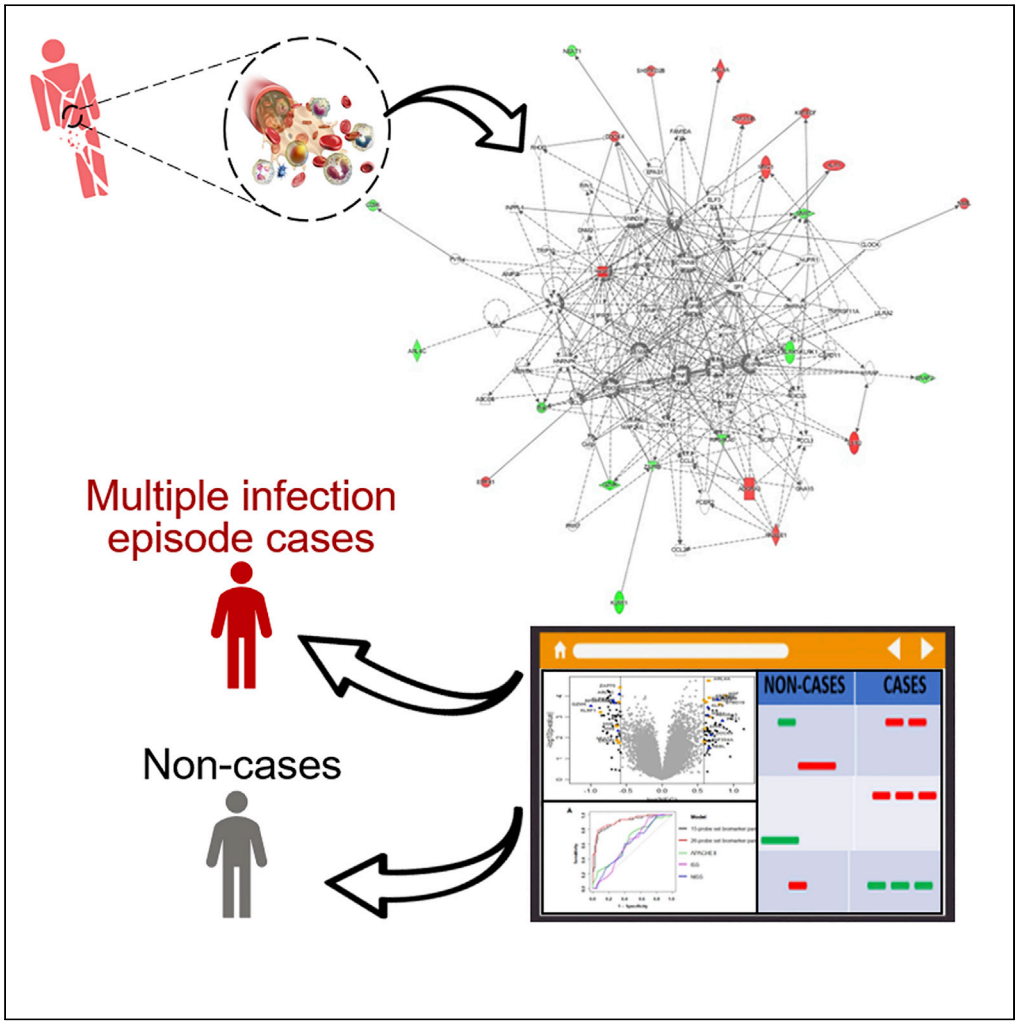
iScience

CellPress
OPEN ACCESS

Article

Multi-Biomarker Prediction Models for Multiple Infection Episodes Following Blunt Trauma



Amy Tsurumi, Patrick J. Flaherty, Yok-Ai Que, ..., Laura F. Goodfield, Ronald G. Tompkins, Laurence G. Rahme

rahme@molbio.mgh.harvard.edu

HIGHLIGHTS

We describe a method for predicting multiple independent infection episodes (MIIEs).

We applied machine learning algorithms to transcriptome data to develop models

The biomarker prediction models significantly outperformed clinical models

External validation in another trauma cohort found evidence of generalizability

Tsurumi et al., iScience *23*, 101659
November 20, 2020
https://doi.org/10.1016/j.isci.2020.101659

iScience

CellPress
OPEN ACCESS

Article

Multi-Biomarker Prediction Models for Multiple Infection Episodes Following Blunt Trauma

Amy Tsurumi,[1,2,3] Patrick J. Flaherty,[4] Yok-Ai Que,[5] Colleen M. Ryan,[1,3] April E. Mendoza,[1] Marianna Almpani,[1,2,3] Arunava Bandyopadhaya,[1,2,3] Asako Ogura,[1,2] Yashoda V. Dhole,[1] Laura F. Goodfield,[1] Ronald G. Tompkins,[1] and Laurence G. Rahme[1,2,3,6,*]

SUMMARY

Severe trauma predisposes patients to multiple independent infection episodes (MIIEs), leading to augmented morbidity and mortality. We developed a method to identify increased MIIE risk before clinical signs appear, which is fundamentally different from existing approaches entailing infections' detection after their establishment. Applying machine learning algorithms to genome-wide transcriptome data from 128 adult blunt trauma patients' (42 MIIE cases and 85 non-cases) leukocytes collected $\leq$48 hr of injury and $\geq$3 days before any infection, we constructed a 15-transcript and a 26-transcript multi-biomarker panel model with the least absolute shrinkage and selection operator (LASSO) and Elastic Net, respectively, which accurately predicted MIIE (Area Under Receiver Operating Characteristics Curve [AUROC] [95% confidence intervals, CI]: 0.90 [0.84–0.96] and 0.92 [0.86–0.96]) and significantly outperformed clinical models. Gene Ontology and network analyses found various pathways to be relevant. External validation found our model to be generalizable. Our unique precision medicine approach can be applied to a wide range of patient populations and outcomes.

INTRODUCTION

The high value of tailored approaches in the care of patients is increasingly appreciated (Chaussabel, 2015; Parikh et al., 2016; Sweeney and Wong, 2016). A method to expedite the timeline of threat detection to before infection happens could yield valuable time for early prophylactic and therapeutic interventions. Moreover, the ability to identify patients at high risk for repeated infections, or infection-related morbidity and mortality, might be considered an important measure to fairly allocate resources such as medication, personal protective equipment, or another high-value scarce intervention (The Commonwealth of Massachusetts, 2020; Institute of Medicine, 2013; University of Pittsburgh, 2020).

Trauma is one of the leading causes of morbidity and mortality worldwide (Heron, 2018; Krug et al., 2000). Severe trauma induces various immune-related responses acutely—it can trigger a state of immunosuppression (Islam et al., 2016; Ward, 2005), prolonged inappropriate immune response (Heffernan et al., 2012; Huber-Lang et al., 2018), leukocytosis (Paladino et al., 2010), and the elevation of specific subpopulations of myeloid cells (Cuenca et al., 2011). Among trauma patients, infections and infections-related complications contribute to substantial mortality and morbidity and prolonged hospital stay, significantly adding to healthcare costs (Cole et al., 2014; Dutton et al., 2010; Glance et al., 2011; Hashmi et al., 2014). Infections and infections-related outcomes vary across individuals, suggesting the importance of considering individual patients' underlying susceptibility and the degree of immunosuppression, or inappropriate immune response experienced.

Methods to identify trauma patients with particularly increased risk of infection could be advantageous for ensuring timely and appropriate delivery of preventative measures (such as early immune-modulating nutrition, microbiome modulation, early mobilization, early removal of lines/tubes, taking all transmission-based precautions), improving surveillance, and promoting antibiotic stewardship to limit the emergence of multidrug resistance bacteria, reduce toxicity to patients, and decrease healthcare costs. Previous studies have

[1]Department of Surgery, Massachusetts General Hospital, and Harvard Medical School, 50 Blossom St., Their 340, Boston, MA 02114, USA

[2]Department of Microbiology, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115, USA

[3]Shriners Hospitals for Children-Boston®, 51 Blossom St., Boston, MA 02114, USA

[4]Department of Mathematics and Statistics, University of Massachusetts at Amherst, Amherst, MA 01003, USA

[5]Department of Intensive Care Medicine, Inselspital, Bern University Hospital, University of Bern, Switzerland, 3010 Bern, Switzerland

[6]Lead Contact

*Correspondence: rahme@molbio.mgh.harvard.edu

https://doi.org/10.1016/j.isci.2020.101659

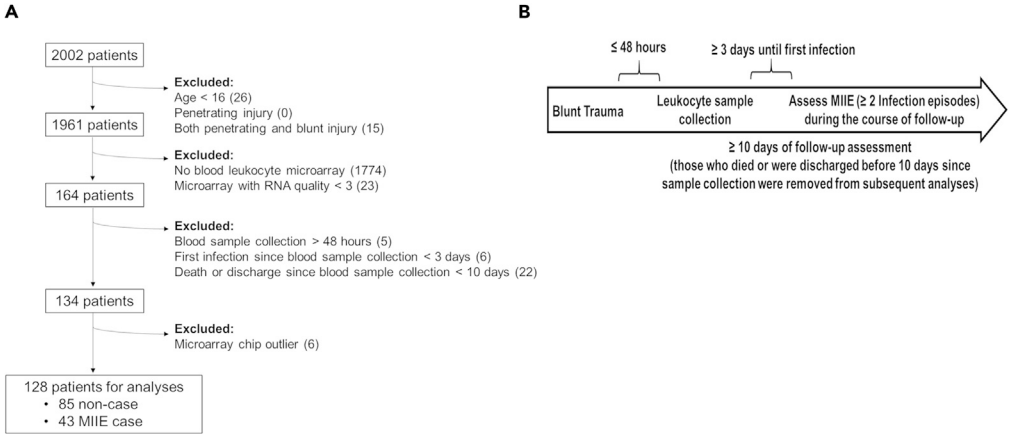


Figure 1. Description of the Patient Population and Study Design

(A and B) (A) Patients who were included/excluded in the study and (B) the study design.

evaluated the use of injury severity scores, such as the Acute Physiology and Chronic Health Evaluation (APACHE) II (Knaus et al., 1985), Injury Severity Score (ISS) (Baker et al., 1974), and New Injury Severity Score (NISS) (Osler et al., 2010) as predictors of infection, in addition to their intended use to predict mortality (Cheadle et al., 1989; Jamulitrat et al., 2002). Using genome-wide transcriptomic information from leukocytes provided at triage to assess the underlying susceptibility, well before the onset of infections, is expected to significantly improve the accuracy of identifying patients who are most at-risk of multiple independent infection episodes. A recent study described that employing a combination of predictors could be more effective than using a single predictor with strong statistical significance, further suggesting that multi-biomarker panels could be highly effective (Lo et al., 2015). Previous studies have utilized transcriptome data in the trauma setting to find transcripts that correlate with poor outcome (Desai et al., 2011) or sepsis (Sweeney et al., 2015). The objective of this study is distinct, as we aim to develop a method to predict multiple infections prior to classic clinical signs of infection. And thus, our approach focuses on the prevention of infections, aiming to predict the outcome before its onset, using early blood samples.

In a previous study among burn trauma patients, we developed a blood transcriptomic multi-biomarker panel for predicting multiple independent infection episode (MIIE) outcomes during the course of recovery (Yan et al., 2015). We employed the least absolute shrinkage and selection operator (LASSO) machine learning algorithm to select probe sets that together (i.e., multi-transcriptome panel) resulted in highly accurate prediction. This model performed significantly better than those based on injury severity assessments at triage and demographic information (Yan et al., 2015). Here, we employed a new approach of combining the use of two algorithms, LASSO and Elastic Net regression, to investigate MIIE outcome. LASSO and Elastic Net were used to reduce the complexity of regression models, in conjunction with cross-validation to select the optimal parameter for reducing the number of predictors. These techniques are highly beneficial in cases such as transcriptome data where the number of potential predictors is extremely large, to overcome the problem of overfitting. LASSO regression reduces highly correlated predictors and selects a minimal panel of predictors, compared to Elastic Net that includes some correlated predictors. We investigated blunt trauma patients in the multi-center Inflammation and Host Response to Injury ("Glue Grant") cohort. This cohort enrolled a high number of patients, generated genome-wide transcriptome data, and collected data longitudinally, allowing us to effectively build a predictive model. Our approach employing unbiased analyses of genome-wide information in the identification of patients at increased risk for MIIE before clinical signs of infection may also be advantageous for providing new insights into the molecular pathways that characterize the pathophysiology underlying hypersusceptibility to infections.

RESULTS

Patient Demographics and Baseline Characteristics

Baseline characteristics of the 128 blunt trauma patients included in the study (Figures 1A and 1B) are presented in Table 1. Motor vehicle collisions were the most frequent injury mechanisms (Table S1). The overall study population consisted of approximately 61.7% males and 38.3% females, with the median age

Variable Description	All Patients (N = 128)	Non-case, ≤ 1 Infection Episode (n = 85)	MIIE Case, ≥2 Infection Episodes (n = 43)	p *value*
Demographic information/clinical descriptors				
Age	34 [25–44]	35 [26–44]	32 [23.5–46]	0.718[a]
Sex				
Female	49 (38.3%)	30 (35.3%)	19 (44.1%)	0.328[c]
Male	79 (61.7%)	58 (64.7%)	23 (55.8%)	
BMI	29.3 ± 7.0	28.8 ± 7.0	30.1 ± 6.9	0.325[b]
BMI categories				
Underweight	1 (0.8%)	0 (0%)	1 (2.3%)	0.336[d]
Healthy	37 (28.9%)	28 (32.9%)	9 (20.9%)	0.216[d]
Overweight	39 (30.5%)	26 (30.6%)	13 (27.9%)	1.000[d]
Obese	51 (39.8%)	31 (36.5%)	20 (46.5%)	0.340[d]
Injury characteristics				
Crush injury	19 (14.8%)	16 (18.8%)	3 (7.0%)	0.113[d]
Severe head injury	13 (10.2%)	8 (9.4%)	5 (11.6%)	0.760[d]
Severity scores				
APACHE II	28 [24–32]	27 [24–31]	29 [27–33.5]	0.008[a]
ISS	34 [22–41.25]	29 [18–41]	36 [27–41.5]	0.041[a]
NISS	37 [27–43.5]	34 [27–43]	41 [30–50]	0.045[a]
AIS (highest of any body part)	4 [3–5]	4 [3–5]	4 [3.5–5]	0.587[a]
Clinical events				
Post-trauma interventions				
Craniotomy	2 (1.6%)	1 (1.2%)	1 (2.3%)	0.402[d]
Thoracotomy	8 (6.3%)	5 (5.9%)	3 (7.0%)	1.000[d]
Laparotomy	61 (47.7%)	35 (41.2%)	26 (60.5%)	0.042[d]
Orthopedic	98 (76.6%)	61 (71.8%)	37 (86.0%)	0.081[d]
Vascular	30 (23.4%)	19 (22.4%)	11 (25.6%)	0.667[d]
Clinical outcomes				
Mortality	5 (3.9%)	3 (3.5%)	2 (4.7%)	1.000[d]
Days in ICU	11 [6–18] *	8 [5–14.75] *	20 [15–29] *	<0.0001[a]
	11.5 [6–18.25]	8 [5–15]	20 [14.5–28]	<0.0001[a]
Discharge day since injury	23 [16.5–33] *	19 [14–26.75] *	35 [27–47] *	<0.0001[a]
	23 [16–33]	19 [13–26]	34 [25.5–47]	<0.0001[a]
Days on ventilator in ICU	8 [4–15]	6 [3–11]	15 [10–23]	<0.0001[a]
Non-infection complications	75 (58.6%)	40 (47.1%)	35 (81.4%)	0.0002[c]
Maximum denver 2 score	2 [1–3]	2 [0–3]	3 [2–3.5]	<0.0001[a]

Table 1. Baseline Characteristics, for the Outcome of Multiple Independent Infection Episodes (MIIEs)

(Continued on next page)

Variable Description	All Patients (N = 128)	Non-case, ≤ 1 Infection Episode (n = 85)	MIIE Case, ≥2 Infection Episodes (n = 43)	p *value*
Maximum marshall score	5.8 [4.0–7.2]	4.7 [3.3–6.4]	6.9 [5.8–8.1]	<0.0001[a]
Maximum central nervous system score	4 [4–4]	4 [4–4]	4 [4–4]	0.064[a]
Maximum cardio score	2.8 [2.0–3.6]	2.4 [1.7–3.2]	3.3 [2.8–4.0]	<0.0001[a]
Maximum respiratory score	2.8 [2.0–3.6]	2.4 [1.7–3.2]	2.8 [2.1–3.1]	<0.0001[a]
Maximum renal score	0.8 [0.7–1.0]	0.8 [0.7–0.9]	0.9 [0.7–1.4]	0.063[a]
Maximum hepatic score	0.0 [0.0–1.2]	0.0 [0.0–0.8]	0.7 [0.0–1.5]	0.002[a]
Maximum hematologic score	0.5 [0.0–1.1]	0.5 [0.0–1.0]	0.6 [0.2–1.1]	0.224[a]

Table 1. Continued

*ICU days and discharge day since injury when calculated only among survivors.

Median [q1-q3], or mean ± SD for continuous variables and n (%) for categorical variables are reported. p values calculations are indicated as follows.

[a]Mann-Whitney U two-tailed test.

[b]unpaired equal variance two-tailed Student's t-test.

[c]Chi-square.

[d]Fisher's Exact two-tailed test.

[interquartile range, IQR] of 34 [25–44] years (Table 1). Common demographic factors were not significantly different between non-cases (≤1 infection episode) and MIIE cases (≥2 infection episodes). However, baseline injury severity scores were significantly higher for MIIE cases (APACHE II score: 27 [24–31] vs. 29 [26–33.5], p = 0.008; ISS: 29 [18–41] vs. 36 [27–41.5], p = 0.041; NISS: 34 [27–43] vs. 41 [30–50], p = 0.045, respectively), while automatic identification system (AIS) (highest score in any body region) was comparable.

As expected, orthopedic procedures were the most frequent surgical interventions that patients received overall (76.6%), followed by laparotomy (47.7%), vascular procedures (23.4%), thoracotomy (6.3%), and craniotomy (1.6%) (Table 1). Apart from the proportion of patients having undergone laparotomy, which was significantly higher among MIIE cases than among non-cases (41.2% vs. 60.5%, p = 0.04), other procedures were similar between non-cases and MIIE cases.

There were five total patients who did not survive, and the cause of death was different for each (Table S2). Mortality was similar between non-cases (3.5%) and MIIE cases (4.7%, p = 1.00). Among survivors, MIIE cases had significantly longer hospital stay than non-cases (discharge at day 19 [14–26.75] vs. 35 [27–47], p < 0.0001). MIIE cases also had a higher proportion of those experiencing non-infection complications than non-cases (47.1% vs. 81.4%, p = 0.0002). Maximum Denver and Marshall scores were significantly higher for MIIE cases than for non-cases (Denver score: 2 [0–3] vs. 3 [2–3.5], p < 0.0001; Marshall score: 4.7 [3.3–6.4] vs. 6.9 [5.8–8.1], p < 0.0001). Among the sub-categories that together determine the total Marshall Score, maximum scores of the following were significantly higher for MIIE cases than for non-cases: cardio (2.4 [1.7–3.2] vs. 2.8 [2.8–4.0], p < 0.0001), respiratory (2.4 [1.7–3.2] vs. 2.8 [2.1–3.1], p < 0.0001), and hepatic score (0.0 [0.0–0.8] vs. 0.7 [0.0–1.5], p < 0.001). The maximum central nervous system, renal, and hematologic scores were not significantly different.

Characteristics of Infection Episodes

(i). Patient Case Outcomes and Timing of Infection Onset

Among the 128 patients in the study, there were 85 non-cases—42 with no infection and 43 with one infection episode—and 42 were MIIE cases (i.e., ≥2 infection episodes). The median [IQR] day for detection of the first infection episode among those who experienced at least one infection episode (i.e., excluding 42 patients those who had no infection episode, for a total of 86 patients) was 8 [5–12] days (Table 2).

(ii) Incidence of Surgical Site Infections versus Other Nosocomial Infections

Among all 128 patients, 36 (28.1%) experienced surgical site infections, compared to 80 (62.5%) who experienced other nosocomial infections (Table 2). Comparing specific subtypes of nosocomial infections, pneumonia (39.1% overall) was highest, followed by urinary tract infection (18.8%), blood infection (17.2%), pseudomembranous colitis (3.9%), catheter-related bloodstream infection (3.9%), empyema (2.3%), and other unspecified infections (5.5%).

(iii) Microorganism Detection

When comparing the incidence of various microorganisms among non-cases with one infection episode versus MIIE cases, relatively higher proportion was found for MIIE cases specifically for Gram positives of *Staphylococcus aureus* (18.2% vs. 39.5%), *Enterococcus* species (11.4% vs. 25.6%), coagulase-negative *staphylococci* (2.3% vs. 16.3%), and *Streptococcus pneumoniae* and *viridans* (2.3% vs. 4.7% for both). The incidence of the Gram-positive *Clostridium* species was the same for non-cases and MIIE cases (both 7.0%) (Table 2). For Gram-negative bacteria, the incidences of the following microorganisms were higher for MIIE cases than for non-cases: *Enterobacter* species (11.4% vs. 37.2%), *Acinetobacter* species (9.1% vs. 30.2%), *Pseudomonas aeruginosa* (11.4% vs. 16.3%), *Haemophilus influenzae* (4.5% vs. 14.0%), *Bacteroides* species (0% vs. 9.3%), *Klebsiella pneumoniae* (2.3% vs. 7.0%), *Neisseria* (0% vs. 7.0%), *Proteus* (2.3% vs. 4.7%), *Serratia marcescens* (2.3% vs. 4.7%), and Gram-negative, not otherwise specified (NOS) (2.3% vs. 11.6%). The incidence was higher among non-cases than among MIIE cases for *Escherichia coli* (11.4% vs. 4.7%) and *Stenotrophomonas* (4.6% vs. 0%). Fungi incidences were higher among MIIE cases than among non-cases: *Candida* species (9.1% vs. 11.6%) and unspecified fungi (0% vs. 2.3%).

(iv) Timing of Microorganism Detection

The median time to the first day of detection of different microorganisms ranged widely (Figure 2). The Gram positives, *Streptococcus* pneumoniae, and *Streptococcus viridans* were found first (at median day 4 and 5, respectively), followed by various Gram negatives and the Gram-positive, *Clostridium* species, between median day 7 and 9.5. Microorganisms that appeared relatively later (day 11–12) include *Candida* species, *Enterobacter* species, and *Serratia marcescens*, *Stenotrophomonas*, *Enterococcus*, Coagulase-negative *staphylococci*, and *Staphylococcus aureus*.

Gene Ontology Reports of 1.5-Fold Changed Probe Sets

We identified 137 probe sets showing at least 1.5-fold upregulated or downregulated difference in expression levels between non-cases and MIIE cases (Figure 3A) and performed Gene Ontology (GO) analyses to find enriched biological processes and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway terms (Table S3). As expected, terms relevant to various immune response pathways were among the top fold enrichment. They included interleukin-4 secretion, regulation of interferon-gamma secretion, cytolysis, regulation of natural killer cell-mediated cytotoxicity, viral genome replication, cellular defense response, adaptive immune response, humoral immune response, T cell co-stimulation, regulation of immune response, T cell receptor signaling pathway, response to tumor necrosis factor, response to virus, immune response, and innate immune response. Other biological process terms with high fold enrichment were signaling pathways with relevance to cell proliferation and differentiation, including regulation of p38 MAPK kinase, calcium-mediated signaling, regulation of fat cell differentiation, organ regeneration, MAP kinase activity, and phosphatidylinositol 3-kinase (PI3K) signaling. Similarly, KEGG pathway terms with high fold enrichment included those relevant to immune response, including natural killer cell-mediated cytotoxicity, malaria, hematopoietic cell lineage, and T cell receptor signaling pathway. Additionally, regulation of cancer, which may have overlapping signaling pathways with infections-related terms, and related to tissue regeneration, also appeared among the enriched terms. Although according to the false discovery rate (FDR)-adjusted p values, statistical significance in enrichment was detected only for a small number of terms, the fold enrichment ranking consistently points to the relevance of immune response terms, as expected.

Prognostic Biomarker Panel Selection

To identify a multi-biomarker panel that collectively predicts the outcome of MIIEs, we analyzed the 137 differentially regulated probe sets by employing a machine learning pipeline that we previously developed and successfully used in our previous study among burn patients (Yan et al., 2015). In the current study, we further added to our analysis pipeline by utilizing a combination of LASSO and Elastic Net regression methods. The LASSO regression that reduces redundancy in predictor selection would allow for a narrow

Variable Description	All Patients (N = 128)	0 and 1 Infection Episode (Non-cases) (n = 85)	1 Infection Episode Only (Non-cases) (n = 43)	≥2 Infection Episodes (MIIE Case) (n = 43)
Characteristics of infection				
First infection day since injury	8 [5–12][a]	8 [5.5–12][a]	8 [5.5–12]	7 [4–11.5]
Surgical site infections	36 (28.1%)	7 (8.2%)	7 (16.3%)	29 (67.4%)
Superficial incisional	7 (5.5%)	2 (2.4%)	2 (4.7%)	5 (11.6%)
Deep incisional	30 (23.4%)	5 (5.9%)	5 (11.6%)	25 (62.8%)
Other infection sites	80 (62.5%)	37 (43.5%)	37 (86.0%)	43 (100%)
Pneumonia	50 (39.1%)	20 (23.5%)	20 (46.5%)	30 (69.8%)
Ventilation-associated pneumonia	46 (35.9%)	20 (23.5%)	20 (46.5%)	26 (60.5%)
Bloodstream infection	22 (17.2%)	4 (4.7%)	4 (9.3%)	18 (41.9%)
Urinary tract infection	24 (18.8%)	7 (8.2%)	7 (16.3%)	17 (39.5%)
Pseudomembranous colitis	5 (3.9%)	3 (3.5%)	3 (7.0%)	2 (4.7%)
Catheter-related bloodstream infection	5 (3.9%)	1 (1.2%)	1 (2.3%)	4 (9.3%)
Empyema	3 (2.3%)	2 (2.4%)	2 (4.7%)	1 (2.3%)
Other	7 (5.5%)	2 (2.4%)	2 (4.7%)	5 (11.6%)
Organism incidence				
Gram-positive bacteria	51 (39.8%)	20 (23.5%)	20 (46.5%)	31 (72.1%)
Staphylococcus aureus	25 (19.5%)	8 (9.4%)	8 (18.2%)	17 (39.5%)
Enterococcus species	16 (12.5%)	5 (5.9%)	5 (11.4%)	11 (25.6%)
Coagulase-negative staphylococci	8 (6.3%)	1 (1.2%)	1 (2.3%)	7 (16.3%)
Clostridium species	6 (4.7%)	3 (3.5%)	3 (7.0%)	3 (7.0%)
Streptococcus pneumoniae	3 (2.3%)	1 (1.2%)	1 (2.3%)	2 (4.7%)
Streptococcus viridans	3 (2.3%)	1 (1.2%)	1 (2.3%)	2 (4.7%)
Gram positive NOS	3 (2.3%)	1 (1.2%)	1 (2.3%)	2 (4.7%)
Gram-negative bacteria	57 (44.5%)	23 (27.1%)	23 (53.5%)	34 (79.1%)
Enterobacter species	21 (16.4%)	5 (5.9%)	5 (11.4%)	16 (37.2%)
Acinetobacter species	17 (13.3%)	4 (4.7%)	4 (9.1%)	13 (30.2%)
Pseudomonas aeruginosa	12 (9.4%)	5 (5.9%)	5 (11.4%)	7 (16.3)
Haemophilus influenza	8 (6.3%)	2 (2.4%)	2 (4.5%)	6 (14.0%)
Escherichia coli	7 (5.5%)	5 (5.9%)	5 (11.4%)	2 (4.7%)
Bacteroides species	4 (3.1%)	0 (0%)	0 (0%)	4 (9.3%)
Klebsiella pneumoniae	4 (3.1%)	1 (1.2%)	1 (2.3%)	3 (7.0%)
Neisseria	3 (2.3%)	0 (0%)	0 (0%)	3 (7.0%)
Proteus	3 (2.3%)	1 (1.2%)	1 (2.3%)	2 (4.7%)
Serratia marcescens	3 (2.3%)	1 (1.2%)	1 (2.3%)	2 (4.7%)

Table 2. Numbers Represent Individuals Who Have Experienced the Indicated Outcomes

(Continued on next page)

Variable Description	All Patients (N = 128)	0 and 1 Infection Episode (Non-cases) (n = 85)	1 Infection Episode Only (Non-cases) (n = 43)	≥2 Infection Episodes (MIIE Case) (n = 43)
Stenotrophomonas species	2 (1.6%)	2 (2.4%)	2 (4.6%)	0 (0%)
Gram negative NOS	6 (4.7%)	1 (1.2%)	1 (2.3%)	5 (11.6%)
Fungi	10 (7.8%)	4 (4.7%)	4 (9.3%)	6 (14.0%)
Candida species	9 (7.0%)	4 (4.7%)	4 (9.1%)	5 (11.6%)
Fungi NOS	1 (0.8%)	0 (0%)	0 (0%)	1 (2.3%)
Unknown	5 (3.9%)	2 (2.4%)	2 (2.3%)	3 (7.0%)

Table 2. Continued

[a]The 42 non-cases who had 0 infection episode, without recordings for first infection day were omitted for these calculations. Recorded species are indicated or described as not otherwise specified (NOS).

selection of a minimal biomarker panel, which is expected to be more practical. Elastic Net regression that includes correlated predictors would allow for a more comprehensive discovery of additional probe sets that are potentially biologically relevant. With LASSO, 15 probe sets were selected, mostly relevant to immune functions and signaling cascades for cellular proliferation and differentiation (Tables 3 and S4; Figures 3A, S2A, and S2B). Upregulated probe sets in the minimal 15-probe set panel included those associated with immune function terms, *hepatocyte growth factor* (*HGF*) *and kelch repeat and BTB domain containing 7* (*KBTBD7*); signaling cascades, *adenosine A3 receptor* (*ADORA3*), *ADP-ribosylation factor-like GTPase 4A* (*ARL4A*), *epiplakin 1* (*EPPK1*), *zinc finger protein 354A* (*ZNF354A*), *SH3 and PX domains 2B* (*SH3PXD2B*); and those with no function term assigned, *RNase A family 1* (*RNASE1*), and *BTB domain containing 19* (*BTBD19*). Downregulated probe sets included those with immune functions, *zeta chain of T cell receptor associated protein kinase 70kDa* (*ZAP70*), *ER aminopeptidase 2* (*ERAP2*), *CD96 molecule* (*CD96*), *membrane metallo-endopeptidase* (*MME*), and *killer cell lectin-like receptor subfamily F, member 1* (*KLRF1*); and a non-coding transcript, *nuclear paraspeckle assembly transcript 1* (*NEAT1*) (Tables 3 and S4; Figures 3A).

With Elastic Net, a total of 26 probe sets were selected that included the 15 probe sets from LASSO and 11 additional ones (Tables 3 and S4; Figures 3A, S2C, and S2D). The additional upregulated probe sets consisted of those with immune functions, *interleukin 1 receptor, type II* (*IL1R2*) and *mannose receptor, C type 1* (*MRC1*); those involved with signaling including, *importin 11* (*IP O 11*), *dedicator of cytokinesis 4* (*DOCK4*) and *Kruppel-like factor 9* (*KLF9*); as well as *nebulette* (*NEBL*) that is important for muscle filament assembly. The additional downregulated probe sets included a different probe set for *MME*, and other immune function genes, *ribosomal protein S6 kinase, 90kDa, polypeptide 5* (*RPS6KA5*), *killer cell lectin-like receptor, subfamily K, member 1* (*KLRK1*) and *granzyme K* (*GZMK*); and the signaling molecule, *ADP-ribosylation factor-like GTPase 4C* (*ARL4C*) (Table 3; Figures 3A and Table S4).

Pathway Network Suggests the Relevance of Major Pathways

We assessed the molecular network connection among the 26 probe sets that were selected by Elastic Net (Figure 3B). The major nodes with the most extensive edges that were central to the connections consisted of major signaling pathway components that are key regulators of inflammation, mitogenic response, and tissue regeneration. These notable nodes included tumor necrosis factor (TNF), transforming growth factor beta-1 (TGFβ-1), and various chemokine (C-X-C motif) ligand (CXCL) members and chemokine (C-C motif) ligand (CCL) members. TNF and TGFβ-1 are major cytokines that can act both synergistically or antagonistically to each other, depending on the cell context. Further substantiating the involvement of these factors, we found p38 mitogen-activated protein kinase (MAPK), which can act downstream of both TNF-α and TGFβ-1 pathways. The other major nodes, extracellular signal-regulated kinases 1/2 (ERK1/2), and mothers against decapentaplegic homolog 3 (SMAD3) are also key downstream components of the TGFβ pathway. A key downstream transcriptional regulator for the canonical Wnt pathway, β-catenin (CTNNB1), another major pathway known to cross talk with TGFβ, was also found as a major node in this network.

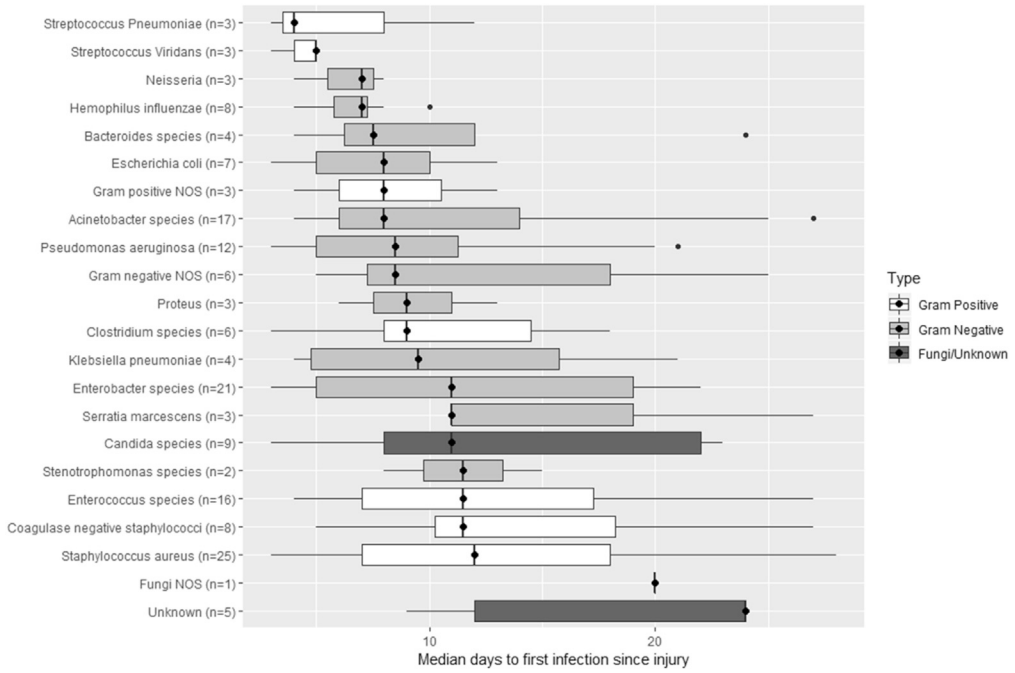


Figure 2. Timing of Detection of Microorganisms

Box plot of the median day of the first infection since the injury, where white boxes indicate Gram positives, light gray boxes indicate Gram negatives, and dark gray boxes indicate fungi and unknown. Recorded species are indicated or described as not otherwise specified (NOS).

Performance in Prediction of MIIEs Using the Multi-Biomarker Panel versus the Clinical Severity Models

The Area Under Receiver Operating Characteristics Curve (AUROC) [95% confidence intervals, CI] of the logistic regression model for predicting MIIE outcomes with the 15-probe set biomarker panel developed with LASSO was 0.90 [0.84–0.96] (Figure 4A). For the 26-probe set panel developed with Elastic Net, there was marginal AUROC increase (0.92 [0.86–0.96], $p = 0.11$) (Figure 4A). Given the objective of the study of developing measures taken early upon arrival at the hospitals for predicting future outcome during recovery, we evaluated common injury severity scores. Compared to the biomarker models, the AUROC of the various injury severity models was notably lower (0.64 [0.54–0.74] for APACHE II, 0.611 [0.51–0.71] for ISS, and 0.61 [0.50–0.71] for NISS, all $p < 0.0001$ compared to the 15-probe set panel model) (Figure 4A). The odds ratios (ORs) [95% CI] for all the covariates of each of the models were also found (Tables S5–S7).

The 15- and 26-probe set panel models had sensitivity [95% CI] of 0.74 [0.59–0.86] vs. 0.79 [0.64–0.90]; specificity [95% CI] of 0.94 [0.87–0.98] vs. 0.94 [0.87–0.98]; positive predictive value (PPV) of 0.86 [0.71–0.95] vs. 0.87 [0.73–0.96]; and negative predictive value (NPV) of 0.88 [0.79–0.94] vs. 0.90 [0.82–0.95], respectively (Table S8). For the various injury severity scores (APACHE II, ISS, NISS), the sensitivity, specificity, PPV, and NPV were generally lower compared to the multi-biomarker panel models (Table S8).

Moreover, we constructed multivariate logistic regression models combining the 15- or the 26-probe set panel with each of the clinical injury severity scores. The AUROC [95% CI] of the 15 probe set panel combined with APACHE II was 0.90 [0.84–0.96], was 0.902 [0.84–0.96] with ISS, and was 0.902 [0.84–0.96] with NISS (Figure 4B). For the 26-probe set panel, the combination with APACHE II was 0.90 [0.86–0.97]; with ISS, it was 0.92 [0.86–0.97]; and with NISS, it was 0.92 [0.86–0.97] (Figure 4B). None of the combined models were significantly different from their respective biomarker only models, suggesting that the addition of clinical score information does not improve the prediction.

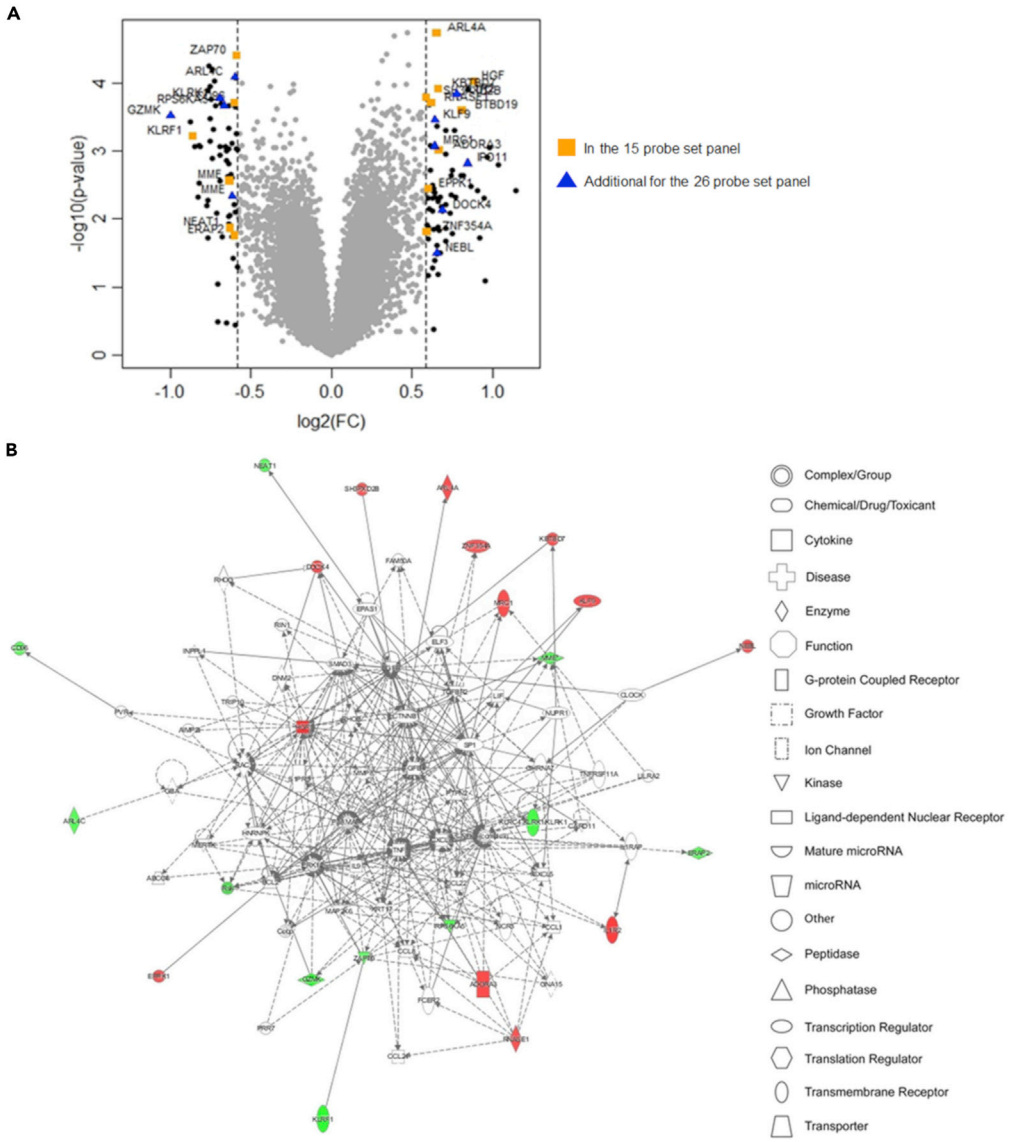


Figure 3. Identification of Probe Sets with 1.5-Fold Change Expression Level Comparing MIIE Cases to Non-cases, and Pathway Analysis

(A) Volcano plot using the initial 25,567 probe sets included in the analyses, where black dotted lines mark the $\log_2(1.5)$ threshold, black dots indicate probe sets with $\geq$1.5-fold change (137 probe sets), orange squares with labels mark data points corresponding to the 15-probe set panel, and blue triangles mark the additional probe sets, that together with the 15 probe sets, result in the 26-probe set panel.

(B) Ingenuity pathway analysis (IPA) of the 26 probe sets that make up the comprehensive biomarker panel. Green marks downregulated and red marks upregulated probe sets. Solid lines indicate known direct interactions, and dotted lines indicate indirect interaction.

External Validation of the Multi-Biomarker Panel in a Different Cohort of Severe Trauma Patients

We performed external validation of our biomarker panel using a severe blunt trauma patient cohort from a previous publication by Cabrera et al. (Cabrera et al., 2017), which had post-admission blood transcriptome $\log_2$ expression values available. The microarray platform used in this study was different from the one used in the Glue Grant, and the measure of one of the genes in the 15-probe set panel was missing. Despite the differences in measurement technology, the lack of measurement for one of the panel, and outcome

Probe Set	Gene Symbol	Gene Name	Fold Change (MIIE/non-cases)	In Both 15- and 26-Probe Set Panel, or Uniquely in 26-Probe Set Panel
210998_s_at	*HGF*	*hepatocyte growth factor*	1.85	15 and 26
238488_at	*IP O 11*	*importin 11*	1.80	26
1557049_at	*BTBD19*	*BTB (POZ) domain containing 19*	1.76	15 and 26
211372_s_at	*IL1R2*	*interleukin 1 receptor, type II*	1.72	26
205003_at	*DOCK4*	*dedicator of cytokinesis 4*	1.61	26
223660_at	*ADORA3*	*adenosine A3 receptor*	1.59	15 and 26
229970_at	*KBTBD7*	*kelch repeat and BTB (POZ) domain containing 7*	1.59	15 and 26
203962_s_at	*NEBL*	*Nebulette*	1.58	26
205020_s_at	*ARL4A*	*ADP-ribosylation factor-like GTPase 4A*	1.57	15 and 26
203543_s_at	*KLF9*	*Kruppel-like factor 9*	1.56	26
204438_at	*MRC1*	*mannose receptor, C type 1*	1.56	26
201785_at	*RNASE1*	*ribonuclease, RNase A family 1*	1.54	15 and 26
232164_s_at	*EPPK1*	*epiplakin 1*	1.52	15 and 26
205427_at	*ZNF354A*	*zinc finger protein 354A*	1.51	15 and 26
231823_s_at	*SH3PXD2B*	*SH3 and PX domains 2B*	1.50	15 and 26
214032_at	*ZAP70*	*zeta chain of T cell receptor-associated protein kinase 70kDa*	0.67	15 and 26
202208_s_at	*ARL4C*	*ADP-ribosylation factor-like GTPase 4C*	0.66	26
227462_at	*ERAP2*	*ER aminopeptidase 2*	0.66	15 and 26
206761_at	*CD96*	*CD96 molecule*	0.66	15 and 26
203434_s_at	*MME*	*membrane metallo-endopeptidase*	0.65	26
238320_at	*NEAT1*	*nuclear paraspeckle assembly transcript 1 (non-protein coding)*	0.65	15 and 26
203435_s_at	*MME*	*membrane metallo-endopeptidase*	0.65	15 and 26
204633_s_at	*RPS6KA5*	*ribosomal protein S6 kinase, 90kDa, polypeptide 5*	0.63	26
1555691_a_at	*KLRK1*	*killer cell lectin-like receptor subfamily K, member 1*	0.62	26
220646_s_at	*KLRF1*	*killer cell lectin-like receptor subfamily F, member 1*	0.55	15 and 26
206666_at	*GZMK*	*granzyme K*	0.50	26

Table 3. Fold Change and P values of the 15-Probe Set Panel from LASSO and 26-Probe Set Panel from Elastic Net, in Order of Magnitude of upregulated to Downregulated

resolution between the Glue Grant and the Cabrera et al. data sets (as described in detail in the methods section), the multi-biomarker panel achieved a relatively high AUROC of 0.76 [0.57–0.96] for the 24-hr post-admission data set and 0.81 [0.62–1.00] for 72-hr post-admission (Figures 5A and 5B). On the other hand, the model with ISS (the only injury severity score that was shared with the Cabrera et al. data set) had a much lower AUROC of 0.64 [0.42–0.86], which was not significantly above 0.5. These results provide evidence for the generalizability of our multi-biomarker panel to different trauma cohort.

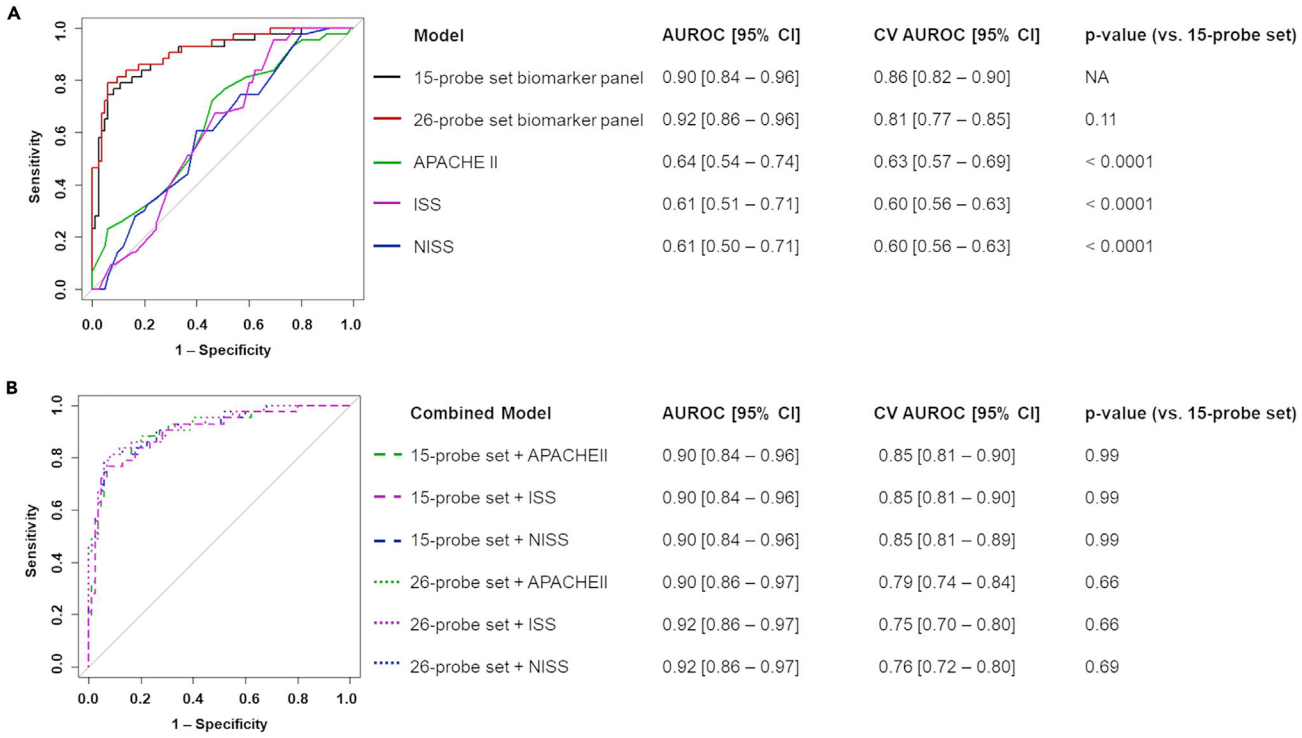


Figure 4. Comparisons of the Performance of the Multi-Biomarker Panel versus the Clinical Severity Models

ROC curves of the various models constructed and the respective AUROC [95% CI], 10-fold cross-validation (CV) AUROC [95% CI], and p value of total AUROC difference compared to the 15-probe set biomarker panel.

(A) Various models: 15-probe set biomarker panel, 26-probe set biomarker panel, APACHE II, ISS, NISS.

(B) Various combined biomarker and clinical score models: 15-probe set biomarker + APACHE, 15-probe set biomarker + ISS, 15-probe set biomarker + NISS, 26-probe set biomarker + APACHE, 26-probe set biomarker + ISS, 26-probe set biomarker + NISS.

DISCUSSION

Our study shows that employing novel prognostic models based on early blood transcriptome profiling following severe trauma is an effective method for identifying patients who are particularly at high risk for MIIE and thus hypersusceptible to infections. That the transcriptome information provides much better prediction than injury severity information argues for the importance of considering each patient's underlying susceptibility and of elucidating relevant molecular mechanisms. The comprehensive data set we used had genome-wide transcriptome and clinical data collected longitudinally from a large number of patients, providing the opportunity to assess early susceptibility. Notably, our results suggest that by measuring the biomarkers among our panel at admission, patients at increased risk for MIIE could be identified before any clinical signs of infection appear. The biomarker panel models in our study had particularly high specificity and NPV measures, while also exerting good sensitivity and PPV. Moreover, when applied to an external validation cohort, it still performed decently. On the other hand, none of the injury severity scores used often at triage in the trauma setting were effective in predicting the MIIE.

The objective of the study was to provide proof-of-concept results for developing a method to gain additional insights into patients' course of recovery from a simple blood draw at admission. Further prospective studies that entail blood draws at admission and measuring the biomarkers we describe here to compare with the severity scores and physiological measures taken normally would provide additional confirmation of the notion that transcriptome data has the potential to improve outcome predictions in the clinical setting. This study focused on the outcome of MIIEs and the potential prevention of infections; however, it is conceivable that the biomarker discovery method described here can be applied to develop prediction methods for other outcomes. A personalized medicine approach and rapid identification of patients with high risk of specific outcomes based on a simple blood draw at admission is expected to improve surveillance, facilitate decision-making and adequate resource allocation, and improve prevention and management before outcomes occur.

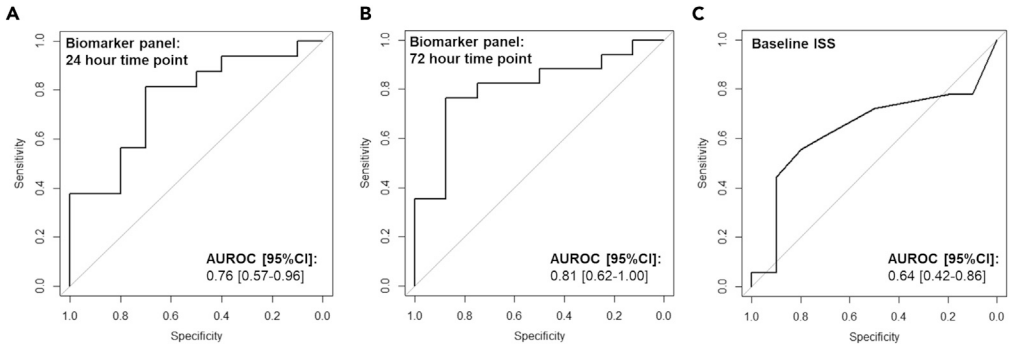


Figure 5. The AUROC Curve When Applying the Biomarker Panel Model Constructed using the Glue Grant, to the Cabrera et al. Data Set as the External Validation Test Set

For both the (A) 24 hr and (B) 72 hr time point of the Cabrera et al. data set, our biomarker panel model conferred significant prediction, as evidenced by the AUROC [95% CI] significantly above 0.5. On the other hand, the prediction model with (C) Injury Severity Score (ISS) did not provide significant prediction.

Our findings could potentially facilitate clinical decision-making by effectively discriminating between those who are expected to develop multiple infections and those who are not. A proposed approach could be that patients who are found not to be hypersusceptible to MIIEs could continue to receive the currently established standard of care, whereas those who are identified to be at high risk could benefit from increased surveillance and additional preventative measures to be taken early. Additional interventions for the high-risk group may include increased surveillance for early mobilization and removal of lines/tubes, coating IV lines and urine catheters with antimicrobials and/or antibiotics, immunomodulatory nutrition therapies (Aghaeepour et al., 2017; Lorenz et al., 2015), and microbiome alterations (Harris et al., 2017; Tosh and McDonald, 2012). Such additional measures would incur unnecessary costs if implemented in all trauma patients; however, it could be cost-effective when used in this targeted set of patients. Efforts aimed at increased prevention have the potential to contribute to alleviating the current antibiotic resistance crisis, toxicity of antibiotics, and the imposed burden on healthcare costs. It is also conceivable that accurate outcome prediction and risk stratification methodologies, such as the one we describe here, could be valuable amid crisis situations that result in severe hospital overload with critically ill patients and scarcity of medical resources. The ability to identify patients at low risk for specific morbidity and mortality could aid in informed prioritization of resource allocation to patients with better potential for recovery and survival (The Commonwealth of Massachusetts, 2020; Institute of Medicine, 2013; University of Pittsburgh, 2020).

Having applied both LASSO and Elastic Net regression methods allowed us to construct a highly predictive model from a minimal set of predictors, meanwhile also more comprehensively assess underlying biological mechanisms by allowing additional transcripts to be included. The LASSO approach selects a stringent set of predictors with less redundancy, which is advantageous in the clinical setting, where a device requiring fewer measurements is more practical and easier to implement. The Elastic Net approach that allows for correlated predictors to be selected found additional transcripts for a more comprehensive discovery of biological mechanisms. The probe sets selected consisted of transcripts with GO terms relevant to infections, as expected, and signaling pertinent to oncogenesis and cancer progression.

In our study, *HGF* was the transcript showing the highest upregulation among MIIE patient blood and in both the 15-probe set and 26-probe set panels. *HGF* and *Met* expression levels have been suggested as a putative biomarker for monitoring infections, as it is well established that the *HGF-Met* signaling pathway deregulation promotes the growth and invasion by various pathogens (Imamura and Matsumoto, 2017). Another upregulated transcript, *ADORA3,* has also been implicated in the clinical setting, and agonists have been developed and shown to induce anti-inflammatory effects by altering the Wnt and NF-κB pathways. As such, the agonists are considered for purposes of treating cancers and inflammatory diseases such as rheumatoid arthritis and psoriasis (Fishman et al., 2012). *NEAT1* is a non-coding RNA that is shown to colocalize with *MALAT1*, a long non-coding RNA often associated with metastatic cancer, at many genomic sites to transcriptionally regulate target genes (West et al., 2014). *CD96* is highly expressed in T and NK cells and well established to be a regulator of immune responses during infection and cancer (Georgiev et al., 2018).

Elastic Net selected two probe sets corresponding to *MME*, providing further support for its importance in MIIE outcome. Studies on its molecular mechanisms and clinical use of inhibitors to its protein product, Neprilysin has been conducted widely, including in Alzheimer's, heart failures, hypertension, and renal diseases (Riddell and Vader, 2017). Our study suggests that its potential role in immunity among patients warrants further investigation. *KLRK1* and *KLRF1*, both killer cell lectin-like receptor subfamily members, were found by Elastic Net, providing evidence of their relevance in infections in the blunt trauma setting. These receptors are abundant on NK cells, and it is well established that they play crucial roles in innate immunity (Barten et al., 2001). These previous findings provide additional confidence in the relevance of our methodology. The pathway analysis found key signaling pathway components among the central nodes having extensive edges, including the major cytokines, TNF, TGFβ-1, CCLs, and CXCLs, as well as key signaling components, p38 MAPK, ERK1/2, SMAD3, and CTNNB1. These components represent the chief signaling pathways that regulate inflammation, mitogenic response, and tissue regeneration, which are also often dysregulated in cancer. Notably, our results suggest that the TNF, TGFβ-1, and Wnt signaling pathways, which are known to also cross talk with one another through downstream cascades, may be important central pathways that explain the interconnection between the prognostic biomarkers identified. These results may suggest that these signaling pathways may represent new host immunomodulatory targets that warrant future mechanistic studies. Follow-up studies in model organisms and controlled studies would aid in establishing whether the genes identified in this study drive susceptibility and in uncovering further mechanistic insights.

It is noteworthy that when comparing the current biomarker panels in the blunt trauma setting with that from our previous study among burn patients (Yan et al., 2015), we observed that none of the transcripts in the panels were shared. These differences may indicate that increased risk depends on the interaction of the type of trauma with each patient's underlying susceptibility to MIIE. Such observation may suggest the need for developing different multi-biomarker panels catered to different types of trauma.

Our study describes methods toward the development of precision medicine tools and offers the possibility of analyses also for outcomes other than multiple infections. The failure of drug trials targeting sepsis (Marshall, 2014; Mitka, 2011) highlights the importance of further studies elucidating the underlying molecular mechanisms and components of heterogeneity in susceptibility to infection and infection-related morbidity within a population. It is conceivable that measuring our biomarker panel to triage patients according to susceptibility to multiple infections will strategically guide prophylactic patient management and help reduce the incidence of infections to limit sepsis (Boomer et al., 2011; Chaussabel, 2015; Parikh et al., 2016). Moreover, the analysis process we describe in this study can potentially also be applied toward biomarker development for sepsis outcome.

This study provides for the first time prediction models for hypersusceptibility to infections, which is highly relevant for critically injured trauma patients, using a machine learning approach. A concern in general for prediction model building is that models may overfit to a specific data set, making them less generalizable. However, using the multi-biomarker panel derived from the Glue Grant population to make predictions in the Cabrera et al. population yielded a relatively high AUROC, demonstrating the broader applicability of our biomarker panel model. These two populations had comparable injury severity; however, they were considerably different in their geographic locations and healthcare systems. Moreover, the gene expression levels were measured by two different transcriptome technologies, and the Cabrera et al. data set was very small in sample size. Despite these differences, our multi-biomarker panels still conferred prediction, providing additional assurance in the validity of our results and evidence for the generalizability of our model. Additional large prospective studies would more rigorously test the validity and generalizability of the multi-biomarker panel identified in this study. Nevertheless, this study provides the first step toward the idea of developing novel approaches for predicting outcomes from blood transcriptome information at admissions.

The value of early MIIE identification, prior to any clinical sign of infection, could be an indispensable tool in other types of trauma and to a wide range of clinical settings. Uncovering biomarkers of increased susceptibility to infections may open new avenues for novel therapeutic targets, as well as contribute to standardizing populations in clinical trials. Although predictive algorithms cannot eliminate medical uncertainty, our analysis method is expected to be widely applicable to other susceptible populations, such as those with diabetes or cardiac disease, the frail elderly population, those treated with immunosuppressive

medication, as well as others. The described methodology of multi-biomarker panel development has the potential to be applied to outcomes and clinical contexts other than MIIE and trauma, providing additional value.

Limitations of the Study

This study entailed a secondary analysis, with external validation using a small data set. Additional external data sets with larger sample sizes, and moreover, a large prospective study would provide additional concrete evidence for the validity and utility of our biomarker panel.

Resource Availability

Lead Contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Laurence G. Rahme (rahme@molbio.mgh.harvard.edu).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

Data set requests should be made to the Glue Grant Consortium, due to human study IRB restrictions. The code for the analyses is available from the lead contact upon request.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101659.

ACKNOWLEDGMENTS

L.G.R. was supported by the Shriners Hospitals Grant #71008 and was partially supported by the Massachusetts General Hospital Research Scholar Award. A.T. was supported by the Shriners Hospitals Research Fellowship #84293. M.A. was supported by the Shriners Hospitals Research Fellowship #84313. A.B. was supported by the Eleanor and Miles Shore Fellowship Program Award for Scholars in Medicine by Harvard Medical School. The investigators acknowledge the contribution of the Inflammation and the Host Response to Injury Large-Scale Collaborative Project Award 5U54GM062119 from the National Institute of General Medical Sciences.

AUTHOR CONTRIBUTIONS

Conceptualization, A.T., P.J.F., Y.Q., C.M.R., L.G.R.; Methodology, A.T., P.J.F., Y.Q., C.M.R., R.G.T., L.G.R.; Software, A.T., P.J.F., L.G.R.; Validation, A.T., P.J.F., L.G.R.; Formal Analysis, A.T.; Resources, R.G.T., L.G.R.; Data Curation, A.T., Y.V.D., L.F.G., L.G.R.; Writing – Original Draft, A.T., P.J.F., Y.Q., C.M.R., L.G.R.; Writing – Review & Editing, A.T., P.J.F., Y.Q., C.M.R., A.E.M., M.A., A.B., A.O., R.G.T., L.G.R.; Visualization, A.T., L.G.R.; Supervision, L.G.R.; Project Administration, R.G.T., L.G.R.; Funding Acquisition, R.G.T., L.G.R.

DECLARATION OF INTERESTS

L.G.R. has a financial interest in Spero Therapeutics, a company developing therapies for the treatment of bacterial infections. L.G.R.'s financial interests were reviewed and are managed by Massachusetts General Hospital and Partners HealthCare in accordance with their conflict of interest policies. The rest of the authors declare no competing interests.

Received: March 27, 2020
Revised: August 25, 2020
Accepted: October 5, 2020
Published: November 20, 2020

REFERENCES

Aghaeepour, N., Kin, C., Ganio, E.A., Jensen, K.P., Gaudilliere, D.K., Tingle, M., Tsai, A., Lancero, H.L., Choisy, B., McNeil, L.S., et al. (2017). Deep immune profiling of an arginine-enriched nutritional intervention in patients undergoing surgery. J. Immunol. ji1700421.

Baker, S.P., O'Neill, B., Haddon, W., Jr., and Long, W.B. (1974). The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. J. Trauma *14*, 187–196.

Barten, R., Torkar, M., Haude, A., Trowsdale, J., and Wilson, M.J. (2001). Divergent and convergent evolution of NK-cell receptors. Trends Immunol. *22*, 52–57.

Boomer, J.S., To, K., Chang, K.C., Takasu, O., Osborne, D.F., Walton, A.H., Bricker, T.L., Jarman, S.D., 2nd, Kreisel, D., Krupnick, A.S., et al. (2011). Immunosuppression in patients who die of sepsis and multiple organ failure. JAMA *306*, 2594–2605.

Cabrera, C.P., Manson, J., Shepherd, J.M., Torrance, H.D., Watson, D., Longhi, M.P., Hoti, M., Patel, M.B., O'Dwyer, M., Nourshargh, S., et al. (2017). Signatures of inflammation and impending multiple organ dysfunction in the hyperacute phase of trauma: a prospective cohort study. Plos Med. *14*, e1002352.

Chaussabel, D. (2015). Assessment of immune status using blood transcriptomics and potential implications for global health. Semin. Immunol. *27*, 58–66.

Cheadle, W.G., Wilson, M., Hershman, M.J., Bergamini, D., Richardson, J.D., and Polk, H.C., Jr. (1989). Comparison of trauma assessment scores and their use in prediction of infection and death. Ann. Surg. *209*, 541–545.

Cole, E., Davenport, R., Willett, K., and Brohi, K. (2014). The burden of infection in severely injured trauma patients and the relationship with admission shock severity. J. Trauma Acute Care Surg. *76*, 730–735.

Cuenca, A.G., Delano, M.J., Kelly-Scumpia, K.M., Moreno, C., Scumpia, P.O., Laface, D.M., Heyworth, P.G., Efron, P.A., and Moldawer, L.L. (2011). A paradoxical role for myeloid-derived suppressor cells in sepsis and trauma. Mol. Med. *17*, 281–292.

Desai, K.H., Tan, C.S., Leek, J.T., Maier, R.V., Tompkins, R.G., Storey, J.D., and Inflammation, and the Host Response to Injury Large-Scale Collaborative Research Program. (2011). Dissecting inflammatory complications in critically injured patients by within-patient gene expression changes: a longitudinal clinical genomics study. PLoS Med. *8*, e1001093.

Dutton, R.P., Stansbury, L.G., Leone, S., Kramer, E., Hess, J.R., and Scalea, T.M. (2010). Trauma mortality in mature trauma systems: are we doing better? An analysis of trauma mortality patterns, 1997-2008. J. Trauma *69*, 620–626.

Fishman, P., Bar-Yehuda, S., Liang, B.T., and Jacobson, K.A. (2012). Pharmacological and therapeutic effects of A3 adenosine receptor agonists. Drug Discov. Today *17*, 359–366.

Georgiev, H., Ravens, I., Papadogianni, G., and Bernhardt, G. (2018). Coming of age: CD96 emerges as modulator of immune responses. Front. Immunol. *9*, 1072.

Glance, L.G., Stone, P.W., Mukamel, D.B., and Dick, A.W. (2011). Increases in mortality, length of stay, and cost associated with hospital-acquired infections in trauma patients. Arch. Surg. *146*, 794–801.

Harris, V.C., Haak, B.W., Boele van Hensbroek, M., and Wiersinga, W.J. (2017). The intestinal microbiome in infectious diseases: the clinical relevance of a rapidly emerging field. Open Forum Infect. Dis. *4*, ofx144.

Hashmi, A., Ibrahim-Zada, I., Rhee, P., Aziz, H., Fain, M.J., Friese, R.S., and Joseph, B. (2014). Predictors of mortality in geriatric trauma patients: a systematic review and meta-analysis. J. Trauma Acute Care Surg. *76*, 894–901.

Heffernan, D.S., Monaghan, S.F., Thakkar, R.K., Machan, J.T., Cioffi, W.G., and Ayala, A. (2012). Failure to normalize lymphopenia following trauma is associated with increased mortality, independent of the leukocytosis pattern. Crit. Care *16*, R12.

Heron, M. (2018). Deaths: leading causes for 2016. Natl. Vital Stat. Rep. *67*, 1–77.

Huber-Lang, M., Lambris, J.D., and Ward, P.A. (2018). Innate immune responses to trauma. Nat. Immunol. *19*, 327–341.

Imamura, R., and Matsumoto, K. (2017). Hepatocyte growth factor in physiology and infectious diseases. Cytokine *98*, 97–106.

Institute of Medicine (2013). Crisis Standards of Care: A Toolkit for Indicators and Triggers (The National Academies Press).

Islam, M.N., Bradley, B.A., and Ceredig, R. (2016). Sterile post-traumatic immunosuppression. Clin. Transl Immunol. *5*, e77.

Jamulitrat, S., Narong, M.N., and Thongpiyapoom, S. (2002). Trauma severity scoring systems as predictors of nosocomial infection. Infect. Control Hosp. Epidemiol. *23*, 268–273.

Knaus, W.A., Draper, E.A., Wagner, D.P., and Zimmerman, J.E. (1985). Apache II: a severity of disease classification system. Crit. Care Med. *13*, 818–829.

Krug, E.G., Sharma, G.K., and Lozano, R. (2000). The global burden of injuries. Am. J. Public Health *90*, 523–526.

Lo, A., Chernoff, H., Zheng, T., and Lo, S.H. (2015). Why significant variables aren't automatically good predictors. Proc. Natl. Acad. Sci. U S A. *112*, 13892–13897.

Lorenz, K.J., Schallert, R., and Daniel, V. (2015). Immunonutrition - the influence of early postoperative glutamine supplementation in enteral/parenteral nutrition on immune response, wound healing and length of hospital stay in multiple trauma patients and patients after extensive surgery. GMS Interdiscip. Plast. Reconstr. Surg. DGPW *4*, Doc15.

Marshall, J.C. (2014). Why have clinical trials in sepsis failed? Trends Mol. Med. *20*, 195–203.

Mitka, M. (2011). Drug for severe sepsis is withdrawn from market, fails to reduce mortality. JAMA *306*, 2439–2440.

Osler, T., Glance, L.G., and Hosmer, D.W. (2010). Simplified estimates of the probability of death after burn injuries: extending and updating the baux score. J. Trauma *68*, 690–697.

Paladino, L., Subramanian, R.A., Bonilla, E., and Sinert, R.H. (2010). Leukocytosis as prognostic indicator of major injury. West J. Emerg. Med. *11*, 450–455.

Parikh, R.B., Kakad, M., and Bates, D.W. (2016). Integrating predictive analytics into high-value care: the dawn of precision delivery. JAMA *315*, 651–652.

Riddell, E., and Vader, J.M. (2017). Potential expanded indications for Neprilysin inhibitors. Curr. Heart Fail. Rep. *14*, 134–145.

Sweeney, T.E., Shidham, A., Wong, H.R., and Khatri, P. (2015). A comprehensive time-course-based multicohort analysis of sepsis and sterile inflammation reveals a robust diagnostic gene set. Sci. Transl Med. *7*, 287ra271.

Sweeney, T.E., and Wong, H.R. (2016). Risk stratification and prognosis in sepsis: what have we learned from microarrays? Clin. Chest Med. *37*, 209–218.

The Commonwealth of Massachusetts (2020). Crisis Standards of Care Planning Guidance for the COVID-19 Pandemic (Department of Public Health, Executive Office of Health and Human Services).

Tosh, P.K., and McDonald, L.C. (2012). Infection control in the multidrug-resistant era: tending the human microbiome. Clin. Infect. Dis. *54*, 707–713.

University of Pittsburgh (2020). A Model Hospital Policy for Allocating Scarce Critical Care Resources (School of Medicine, Department of Critical Care Medicine).

Ward, P.A. (2005). Immunosuppression after trauma. Crit. Care Med. *33*, 1453–1454.

West, J.A., Davis, C.P., Sunwoo, H., Simon, M.D., Sadreyev, R.I., Wang, P.I., Tolstorukov, M.Y., and Kingston, R.E. (2014). The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. Mol. Cell *55*, 791–802.

Yan, S., Tsurumi, A., Que, Y.A., Ryan, C.M., Bandyopadhaya, A., Morgan, A.A., Flaherty, P.J., Tompkins, R.G., and Rahme, L.G. (2015). Prediction of multiple infections after severe burn trauma: a prospective cohort study. Ann. Surg. *261*, 781–792.

iScience, Volume *23*

Supplemental Information

Multi-Biomarker Prediction Models for Multiple Infection Episodes Following Blunt Trauma

Amy Tsurumi, Patrick J. Flaherty, Yok-Ai Que, Colleen M. Ryan, April E. Mendoza, Marianna Almpani, Arunava Bandyopadhaya, Asako Ogura, Yashoda V. Dhole, Laura F. Goodfield, Ronald G. Tompkins, and Laurence G. Rahme

TRANSPARENT METHODS

Study design/patient

We performed a secondary analysis of patient clinical and genomic data from the Glue Grant, a prospective, longitudinal study that enrolled patients at seven US Level 1 trauma centers between 2003 and 2009. Permission for the use of de-identified data was obtained from the Massachusetts General Hospital Institutional Review Board (MGH IRB protocol 2008-P-000629/1). Among the 2,002 patients in the dataset, our inclusion criteria identified 128 adult (age ≥ 16 years) patients who suffered blunt trauma (excluding penetrating injury only, or blunt with penetrating injury), have follow-up time of at least 10 days since blood collection, and having early (≤ 48 hours since trauma injury) blood microarray transcriptome data of high RNA quality (RNA quality ≥ 3 out of 4) and chips that were not outliers. At least three days from time of blood draw to first recorded infection (Figure 1A, B). Where patients had multiple blood samples that fit the criteria, the earliest was used. Clinical outcomes and injury severity scores (APACHEII, ISS, NISS, Denver 2, Marshall) were recorded by participating institutions according to the guidelines outlined by the Glue Grant Consortium. Body mass index (BMI) was calculated from recorded height and weight, and BMI categories were assigned according to the Centers for Disease Control and Prevention's recommendations. For assigning categories of major clinical procedures, recorded sub-categories were consolidated to create more general categories. Similarly, various non-infection complications and surgical site infection sub-categories were consolidated, as indicated below.

Major clinical procedures categories:

For assigning categories of major clinical procedures, sub-categories were combined as follows: "Laparotomy" includes entries of laparotomy NOS, and laparotomy with other procedure, with splenectomy, with repair/packing of the liver, with the repair of gastric, duodenal or small bowel perforation, with the repair of the large bowel or rectal perforation, with nephrectomy, with repair

of major vascular injury, with drainage of intra-abdominal abscess, and second-look laparotomy or abdominal washout. "Orthopedic" includes soft tissue debridement/amputations, internal fixation of femur, open skeletal fixation exclusive of femur, percutaneous skeletal fixation, and spine fixation. "Vascular" includes peripheral vascular and angiographic embolization. "Craniotomy" and "Thoracotomy" were used as recorded in the original dataset.

Surgical site infections categories:

For surgical site infection categories, "Superficial incisional infection" includes recordings of the shoulder girdle, lumbar spine (bony), lower extremity, and abdomen and pelvis (non-bony) indicated as superficial incisional and "Deep incisional" combines upper extremity, lower extremity, pelvis (body), and abdomen and pelvis (non-body) recorded as deep incisional, as well as chest (pleural space), abdomen and pelvis (non-bony), upper extremity, and lower extremity recorded as organ/space.

Outcome definition:

Patients were assigned to MIIE-cases (≥2 cumulative infection episodes) and non-cases (≤ 1 infection episode), following the same method as in our previous study (Yan et al., 2015), in which a decision tree considering the timing and type of infection and the isolated pathogen was used to tabulate total independent infection episodes (Supplementary Figure S1).

QUANTIFICATION AND STATISTICAL ANALYSIS

Software used

R version 3.4.4 was used for the statistical analyses, as described below, with the following packages and versions: GCRMA 2.50.0 (Wu and Irizarry, 2017), arrayQualityMetrics 3.34.0 (Kauffmann et al., 2009), EMA 1.4.5 (Servant et al., 2010), LIMMA 3.34.9 (Ritchie et al., 2015), Glmnet 2.0-16 (Friedman et al., 2010), pROC 1.13.0 (Robin et al., 2011), Caret 6.0-81 (Kuhn, 2008), epiR 0.9-99 (Stevenson et al., 2019).

Statistical Analyses

Baseline characteristics are reported as medians with interquartile range, means with standard deviation, or total numbers with proportions in percentages, as indicated in the legend (Table 1). Medians between two groups (non-cases versus MIIE-cases) were compared using the Mann-Whitney U test, and means were analyzed by the unpaired t-test assuming equal variance. For comparing proportions, the Chi-square test was used for expected values 5 or greater, or Fisher's exact test was used for an expected value less than 5.

For processing microarray data files, first, the GCRMA package was employed to obtain normalized $\log_2$ expression values of probe sets. Then chips that were flagged as outliers by the arrayQualityMetrics package were excluded. Subsequently, internal control probe sets were removed, and the EMA package was used to filter out low abundance probe sets (below the threshold of 3.5 $\log_2$ expression value across all samples). These filtering steps reduced the number of probe sets from 54,675 to 25,567 for subsequent analyses. The LIMMA package was used to identify probe sets with at least 1.5-fold difference between the non-cases and MIIE-cases; this reduced the number of probe sets to 137. Functional annotation analyses were conducted using Database for Annotation, Visualization, and Integrated Discovery (DAVID), version 6.8 (Huang da et al., 2009), using the 1.5-fold changed 137 probe sets as the target and total 25,567 probe sets as the background set. The fold enrichment, unadjusted and FDR-adjusted p-values were reported.

The Glmnet package was used to implement the Least absolute shrinkage and selection operator (LASSO) regression and Elastic Net regression, to select probe sets that were predictive of MIIE, as previously conducted (Yan et al., 2015). The penalty weight, lambda (λ), was identified by finding the 10-fold cross-validation (CV) error, repeated 100 times. Probe sets were selected according to the value of λ that yielded minimum average binomial deviance plus one standard error on the test set (λ_{1se}), rather than minimum (λ_{min}) to limit overfitting. For LASSO, the λ_{1se} was found to be 0.068, and for the Elastic net, it was found to be 0.108

(Supplementary Figure S2). This biomarker panel selection process yielded15 probe sets with LASSO and 26 probe sets with Elastic net. For building the clinical model, similarly to our previous study (Yan et al., 2015), stepwise regression was conducted to select each of the severity scores (APACHEII, ISS, NISS), with potential other covariates. For APACHEII, which includes age as one of the factors for the score calculation, sex and race were used for stepwise selection. ISS and NISS were each tested with age, sex, and race. For all the severity scores, univariate models with the scores only were selected.

The following logistic regression models were constructed for the binary MIIE outcome: (a) multivariate model with the 15 probe set LASSO predictors, (b) multivariate model with the 26 probe sets Elastic Net predictors, (c) univariate models of various clinical scores (APACHEII ISS, NISS), and (d) multivariate models with a combination of the 15, or 26 probe sets with each of the clinical scores. The area under the receiver operating characteristic curve (AUROC) with bootstrap 95% confidence intervals were estimated and used to compare the models using the pROC package. The comparison was repeated for 10-fold cross-validation sampling using the Caret package. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) and confidence intervals were calculated using the epiR package, using the optimal probability cut-off determined as the top-left corner of the ROC curve for each model.

For the volcano plot, $\log_2$ fold change comparing MIIE-cases and non-cases was plotted on the x-axis and p-values on the y-axis, of all initial 25,567 probe sets after the filtering step. Data points of probe sets with at least 1.5-fold expression level difference (137 probe sets) are marked with black dots. Among these 137 probe sets, those corresponding to the 15 probe set panel selected with LASSO are marked with orange squares, and additional probe sets selected by Elastic Net for the 26 probe set panel are marked with blue triangles. Ingenuity Pathway Analyses (IPA) was used to generate an interaction network of the 26 probe sets selected by Elastic Net (QIAGEN) (Kramer et al., 2014).

External Validation Dataset and Statistical Analyses

Blood transcriptome data for external validation was obtained from the Cabrera *et al.,* which conducted a secondary analysis of the Activation of Coagulation and Inflammation in Trauma (ACIT2) cohort (Cabrera et al., 2017). The ACIT2 study enrolled adult trauma patients at the Royal London Hospital, and the transcriptome study was conducted among patients enrolled between 2008-2012. The datasets with normalized Illumina microarray $\log_2$ expression levels, accompanying clinical information, and infections outcome were downloaded from: https://github.com/C4TS/HyperacutePhase. For our analysis, we included 28 unique critically injured trauma patients with baseline ISS and infections outcome information available, after removing those with missing information. The Cabrera *et al.* data set does not have the resolution of the infection outcome that our Glue grant data set does; therefore we classified infection "yes" or "no" status, as indicated in their dataset. This subpopulation's ISS range was 25-51 (median 31), and the age range was 17-69 (median 37.5), which were comparable to the Glue Grant population. Among them, 26 had blood transcriptome data available for the 24 hour time point, and 25 had data available for the 72 hour time. Expression values of all the genes represented in the 15 probe set panel, except *SH3PXD2B*, was available in the Cabrera *et al.* dataset ("NormData" file). The data processing steps in Cabrera *et al.* entailed filtering out "nonexpressed" probe sets that did not pass the detection p-value threshold of 0.05 in at least three arrays. Therefore, we assumed the expression values for *SH3PXD2B* to be 0 for our analysis. We applied the 15 probe set panel model to calculate the predicted probabilities for the subjects in the Cabrera *et al.* dataset and constructed the ROC curve.

Supplementary Tables and Figures

Injury mechanisms				
Fall	7 (5.5%)	6 (7.1%)	1 (2.3%)	0.42[d]
Machinery	1 (0.8%)	1 (1.2%)	0 (0%)	1.00[d]
MVC – occupant	70 (54.7%)	43 (50.6%)	27 (62.8%)	0.26[d]
MVC – motorcyclist	23 (18.0%)	16 (18.8%)	7 (16.3%)	0.81[d]
MVC – cyclist	3 (2.3%)	3 (3.5%)	0 (0%)	0.55[d]
MVC – pedestrian	19 (14.8%)	13 (15.3%)	6 (14.0%)	1.00[d]
Assault	1 (0.8%)	1 (1.2%)	0 (0%)	1.00[d]
Other	4 (3.1%)	2 (2.4%)	2 (4.7%)	0.60[d]

Supplementary Table S1. Injury mechanisms among all 128 patients in the study, related to Table 1.

Death day since injury	**Primary cause of death**	**Secondary cause of death**	**MIIE case**
10	Multiple organ failure		No
10	Sepsis	Multiple organ failure	No
20	Hypoxia		No
11	Severe head injury trauma	Cardiac dysfunction	Yes
24	Brain death		Yes

Supplementary Table S2. Causes of death among the 5 non-survivors in the study, related to Table 1.

Clinical models	**OR [95% CI]**	***p-value***
APACHE II	1.11 [1.03 – 1.19]	<0.01 **
ISS	1.03 [1.00 – 1.06]	0.06
NISS	1.03 [1.00 – 1.06]	0.06

Supplementary Table S5. OR estimates [95%CI] and coefficient p-values of the clinical severity score univariate models, related to Fig. 4.

15 probe set biomarker panel multivariate model:			**26 probe set biomarker panel multivariate model:**		
MODEL	**OR [95% CI]**	***p-value***	**MODEL**	**OR [95% CI]**	***p-value***
205427_at: ZNF354A	2.20 [1.42 – 3.81]	< 0.01**	202208_s_at: ARL4C	4.12 [0.47 – 44.23]	0.21
232164_s_at: EPPK1	1.94 [1.14 – 3.50]	0.02*	1555691_a_at: KLRK1	3.70 [0.97 – 17.46]	0.07
201785_at: RNASE1	1.78 [0.90 – 3.82]	0.11	231823_s_at: SH3PXD2B	2.58 [0.65 – 13.11]	0.21
1557049_at: BTBD19	1.69 [0.82 – 3.69]	0.17	232164_s_at: EPPK1	2.30 [1.21 – 4.86]	0.02*
223660_at: ADORA3	1.61 [0.81 – 3.48]	0.20	205427_at: ZNF354A	2.27 [1.30 – 4.45]	< 0.01**
229970_at: KBTBD7	1.43 [0.65 – 3.26]	0.38	201785_at: RNASE1	2.12 [0.92 – 5.56]	0.09
231823_s_at: SH3PXD2B	1.16 [0.45 – 3.17]	0.77	203434_s_at: MME	2.10 [0.22 – 25.75]	0.54
206761_at: CD96	0.94 [0.23 – 3.55]	0.93	1557049_at: BTBD19	1.97 [0.76 – 5.25]	0.16
205020_s_at: ARL4A	0.90 [0.28 – 2.80]	0.86	204438_at: MRC1	1.92 [0.77 – 5.29]	0.18
214032_at: ZAP70	0.81 [0.22 – 3.13]	0.76	223660_at: ADORA3	1.76 [0.75 – 4.59]	0.21
227462_at: ERAP2	0.67 [0.41 – 1.03]	0.08	203962_s_at: NEBL	1.45 [0.94 – 2.29]	0.10
210998_s_at: HGF	0.65 [0.29 – 1.35]	0.26	229970_at: KBTBD7	1.43 [0.42 – 5.17]	0.57
220646_s_at: KLRF1	0.61 [0.34 – 1.06]	0.09	238488_at: IPO11	1.24 [0.58 – 2.76]	0.58
238320_at: NEAT1	0.53 [0.32 – 0.83]	< 0.01**	204633_s_at: RPS6KA5	1.17 [0.36 – 3.93]	0.79
203435_s_at: MME	0.50 [0.25 – 0.91]	0.04*	203543_s_at: KLF9	1.01 [0.29 – 3.57]	0.99
			205003_at: DOCK4	0.96 [0.55 – 1.75]	0.90
			211372_s_at: IL1R2	0.85 [0.20 – 3.55]	0.83
			206761_at: CD96	0.80 [0.12 – 5.05]	0.81
			205020_s_at: ARL4A	0.60 [0.14 – 2.30]	0.47
			227462_at: ERAP2	0.58 [0.32 – 0.97]	< 0.05*
			206666_at: GZMK	0.57 [0.21 – 1.48]	0.26
			238320_at: NEAT1	0.39 [0.18 – 0.73]	< 0.01**
			220646_s_at: KLRF1	0.35 [0.14 – 0.82]	0.02*
			210998_s_at: HGF	0.33 [0.09 – 0.99]	0.07
			203435_s_at: MME	0.31 [0.03 – 2.38]	0.28
			214032_at: ZAP70	0.28 [0.03 – 2.04]	0.22

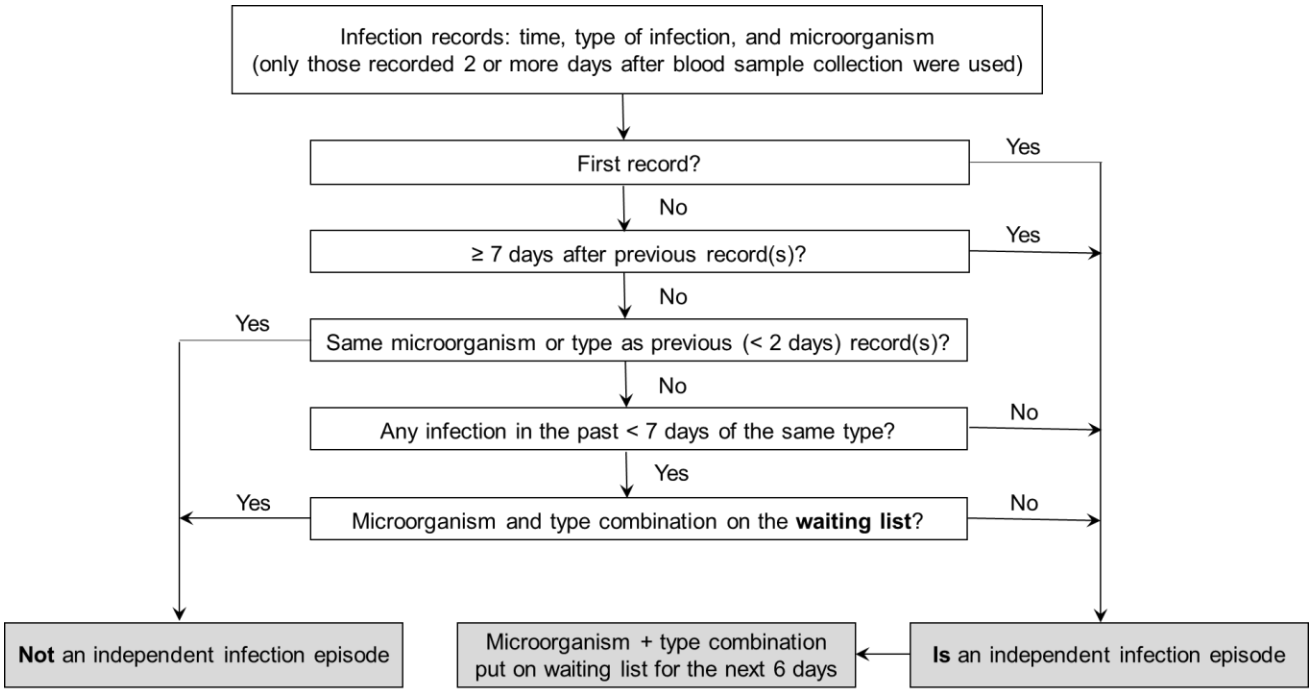
Supplementary Table S6. OR estimates [95%CI] and coefficient p-values of the biomarker panel multivariate models, related to Fig. 4.

MODEL	+ APACHE II		+ ISS		+ NISS	
	OR [95% CI]	***p-value***	**OR [95% CI]**	***p-value***	**OR [95% CI]**	***p-value***
202208_s_at: ARL4C	3.89 [0.45 – 41.33]	0.23	3.74 [0.40 – 45.88]	0.27	4.32 [0.45 – 54.91]	0.23
1555691_a_at: KLRK1	3.76 [0.99 – 17.66]	0.07	3.66 [0.96 – 17.26]	0.07	3.74 [0.97 – 17.79]	0.07
231823_s_at: SH3PXD2B	2.64 [0.65 – 13.84]	0.21	2.50 [0.63 – 12.95]	0.23	2.62 [0.65 – 13.77]	0.21
232164_s_at: EPPK1	2.30 [1.20 – 4.84]	0.02*	2.26 [1.18 – 4.85]	0.02*	2.32 [1.21 – 4.95]	0.02*
201785_at: RNASE1	2.20 [0.94 – 5.73]	0.08	2.10 [0.92 – 5.50]	0.10	2.28 [1.30 – 4.50]	< 0.01**
205427_at: ZNF354A	2.18 [1.22 – 4.34]	0.01*	2.26 [1.29 – 4.42]	< 0.01**	2.13 [0.92 – 5.69]	0.09
1557049_at: BTBD19	1.98 [0.75 – 5.41]	0.17	1.91 [0.72 – 5.24]	0.20	2.11 [0.22 – 25.96]	0.53
204438_at: MRC1	1.92 [0.77 – 5.33]	0.18	1.91 [0.76 – 5.28]	0.18	2.00 [0.76 – 5.47]	0.16
203434_s_at: MME	1.79 [0.17 – 23.44]	0.64	2.08 [0.22 – 25.72]	0.54	1.93 [0.77 – 5.35]	0.18
223660_at: ADORA3	1.77 [0.75 – 4.72]	0.21	1.71 [0.71 – 4.56]	0.25	1.78 [0.74 – 4.78]	0.21
203962_s_at: NEBL	1.45 [0.94 – 2.30]	0.10	1.44 [0.93 – 2.28]	0.11	1.45 [0.94 – 2.30]	0.10
229970_at: KBTBD7	1.44 [0.43 – 5.18]	0.56	1.44 [0.43 – 5.24]	0.56	1.42 [0.42 – 5.16]	0.57
204633_s_at: RPS6KA5	1.26 [0.37 – 4.41]	0.71	1.18 [0.36 – 3.96]	0.78	1.24 [0.57 – 2.76]	0.59
238488_at: IPO11	1.25 [0.57 – 2.80]	0.58	1.24 [0.58 – 2.75]	0.57	1.16 [0.35 – 3.92]	0.80
203543_s_at: KLF9	0.97 [0.27 – 3.49]	0.96	0.99 [0.28 – 3.56]	0.99	1.01 [0.29 – 3.56]	0.99
205003_at: DOCK4	0.96 [0.54 – 1.74]	0.89	0.97 [0.55 – 1.77]	0.91	0.96 [0.54 – 1.75]	0.89
211372_s_at: IL1R2	0.88 [0.20 – 3.75]	0.86	0.87 [0.20 – 3.65]	0.85	0.85 [0.20 – 3.54]	0.82
206761_at: CD96	0.83 [0.12 – 5.27]	0.85	0.81 [0.12 – 5.04]	0.82	0.80 [0.11 – 5.06]	0.81
205020_s_at: ARL4A	0.64 [0.14 – 2.64]	0.54	0.61 [0.14 – 2.39]	0.49	0.59 [0.13 – 2.31]	0.46
227462_at: ERAP2	0.59 [0.32 – 0.99]	0.06	0.57 [0.31 – 0.96]	< 0.05*	0.58 [0.32 – 0.97]	0.05
206666_at: GZMK	0.58 [0.21 – 1.51]	0.28	0.60 [0.20 – 1.63]	0.33	0.57 [0.20 – 1.50]	0.26
238320_at: NEAT1	0.39 [0.19 – 0.75]	< 0.01**	0.39 [0.19 – 0.74]	< 0.01**	0.38 [0.18 – 0.74]	< 0.01**
220646_s_at: KLRF1	0.37 [0.14 – 0.86]	0.03*	0.36 [0.14 – 0.86]	0.03*	0.35 [0.13 – 0.84]	0.03*
203435_s_at: MME	0.37 [0.03 – 3.23]	0.39	0.31 [0.03 – 2.47]	0.29	0.33 [0.08 – 1.01]	0.08
210998_s_at: HGF	0.33 [0.09 – 0.98]	0.06	0.35 [0.09 – 1.10]	0.10	0.30 [0.03 – 2.37]	0.28
214032_at: ZAP70	0.28 [0.03 – 2.00]	0.21	0.30 [0.03 – 2.38]	0.26	0.27 [0.03 – 2.14]	0.22
Clinical Score	1.02 [0.92 – 1.14]	0.69	1.01 [0.95 – 1.06]	0.80	1.00 [0.94 – 1.05]	0.90

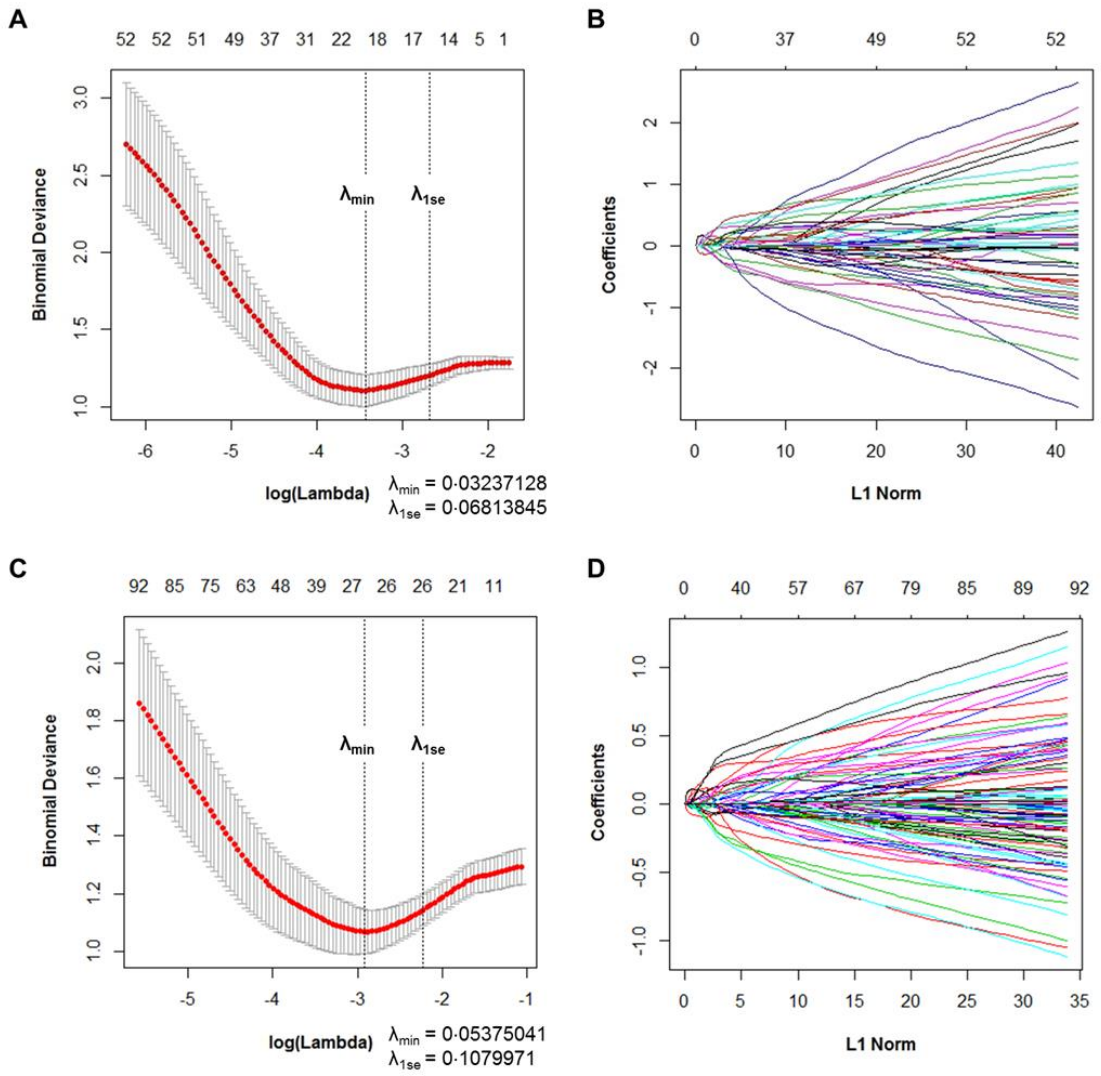
Supplementary Table S7. OR estimates [95%CI] and coefficient p-values of the combined genomic (15 or 26 probe set panel) and clinical severity score multivariate models, related to Fig. 4.

Model	Probability threshold	Corresponding clinical score cut-offs	Sensitivity [95%CI]	Specificity [95%CI]	PPV [95%CI]	NPV [95%CI]
15 biomarker panel	0.561	NA	0.74 [0.59-0.86]	0.94 [0.87-0.98]	0.86 [0.71-0.95]	0.88 [0.79-0.94]
26 biomarker panel	0.476	NA	0.79 [0.64-0.90]	0.94 [0.87-0.98]	0.87 [0.73-0.96]	0.90 [0.82-0.95]
APACHEII	0.315	28.22	0.72 [0.56-0.85]	0.54 [0.43-0.65]	0.44 [0.32-0.57]	0.79 [0.67-0.89]
ISS	0.267	21.02	0.95 [0.84-0.99]	0.31 [0.21-0.42]	0.41 [0.31-0.51]	0.93 [0.76-0.99]
NISS	0.344	39.50	0.60 [0.44-0.75]	0.60 [0.49-0.70]	0.43 [0.31-0.57]	0.75 [0.63-0.85]
15 biomarker panel +APACHEII	0.479	NA	0.77 [0.61-0.88]	0.93 [0.85-0.97]	0.85 [0.69-0.94]	0.89 [0.80-0.94]
15 biomarker panel +ISS	0.511	NA	0.77 [0.61-0.88]	0.93 [0.85-0.97]	0.85 [0.69-0.94]	0.89 [0.80-0.94]
15 biomarker panel +NISS	0.505	NA	0.77 [0.61-0.88]	0.93 [0.85-0.97]	0.85 [0.69-0.94]	0.89 [0.80-0.94]
26 biomarker panel +APACHEII	0.463	NA	0.81 [0.67-0.92]	0.92 [0.84-0.97]	0.83 [0.69-0.93]	0.91 [0.82-0.96]
26 biomarker panel +ISS	0.476	NA	0.79 [0.64-0.90]	0.94 [0.87-0.98]	0.87 [0.73-0.96]	0.90 [0.82-0.95]
26 biomarker panel +NISS	0.446	NA	0.81 [0.67-0.92]	0.92 [0.84-0.97]	0.83 [0.69-0.93]	0.91 [0.82-0.96]

Supplementary Table S8. Sensitivity and specificity, PPV and NPV of the various models, related to Fig. 4**.**



Supplementary Figure S1. Decision tree used to define independent infection episodes from clinical and microbiological records, taking into account, the timing of infections, type of infection and pathogen isolated, related to Fig. 1. Each infection record was placed on a "waiting list" for six days if the types and modes of infection and pathogen were similar, or two days if they were not. This method allowed tabulation of the number of total independent infection episodes that is not an extension of a previously recorded infection from the time of the injury. Adapted from Yan et al, 2015. [25]



Supplementary Figure S2. LASSO and Elastic Net regression outputs, related to Fig. 4. **(A)** A plot of 10-fold CV repetition results, with dotted lines showing λ_{min} and λ_{1se} values, where 15 probe sets were selected. **(B)** LASSO coefficient profile plot of the coefficient paths, where at λ_{1se}, the selected 15 probe sets have coefficients significantly different from zero. **(C)** 10-fold CV results and **(D)** coefficient profile plot for Elastic Net.

SUPPLEMENTARY REFERENCES

Cabrera, C.P., Manson, J., Shepherd, J.M., Torrance, H.D., Watson, D., Longhi, M.P., Hoti, M., Patel, M.B., O'Dwyer, M., Nourshargh, S.*, et al.* (2017). Signatures of inflammation and impending multiple organ dysfunction in the hyperacute phase of trauma: A prospective cohort study. PLoS Med *14*, e1002352.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw *33*, 1-22.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44-57.

Kauffmann, A., Gentleman, R., and Huber, W. (2009). arrayQualityMetrics--a bioconductor package for quality assessment of microarray data. Bioinformatics *25*, 415-416.

Kramer, A., Green, J., Pollard, J., Jr., and Tugendreich, S. (2014). Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics *30*, 523-530.

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. Journal of Statistical Software *28*.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res *43*, e47.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.C., and Muller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics *12*, 77.

Servant, N., Gravier, E., Gestraud, P., Laurent, C., Paccard, C., Biton, A., Brito, I., Mandel, J., Asselain, B., Barillot, E.*, et al.* (2010). EMA - A R package for Easy Microarray data analysis. BMC Res Notes *3*, 277.

Stevenson, M., Nunes, E., Heuer, C., Marshall, J., Sanchez, J., Thornton, R., Reiczigel, J., Robison-Cox, J., Sebastiani, P., Solymos, P.*, et al.* (2019). epiR: Tools for the Analysis of Epidemiological Data. https://CRANR-projectorg/package=epiR.

Wu, J., and Irizarry, R.A. (2017). gcrma: Background Adjustment Using Sequence Information. R package version 2500.

Yan, S., Tsurumi, A., Que, Y.A., Ryan, C.M., Bandyopadhaya, A., Morgan, A.A., Flaherty, P.J., Tompkins, R.G., and Rahme, L.G. (2015). Prediction of multiple infections after severe burn trauma: a prospective cohort study. Ann Surg *261*, 781-792.